

# Repositórios de dados de pesquisa

## Curso 2

A trilha **Ciência Aberta** abrange apostilas criadas pelo grupo RDPBrasil para treinamentos promovidos pela Rede Nacional de Pesquisa (RNP). O curso 2 intitula-se **Repositórios de dados de pesquisa**.

#### **Pesquisadores**

Samile Andréa de Souza Vanz  
Caterina Groposo Pavão  
Rafael Port da Rocha  
Rene Faustino Gabriel Junior  
Sônia Elisa Caregnato

#### **Bolsistas**

Ana Júlia Lopes  
Lucia Helena Cunha Vidal  
Wagner Silva Wessfl

#### **Contato**

site: [www.ufrgs.br/rdpbrasil](http://www.ufrgs.br/rdpbrasil)  
Rua Ramiro Barcelos, 2705 - Campus Saúde  
Sala: 106 - Anexo 1  
Brasil - Porto Alegre - RS - CEP 90.035-007  
Telefone: +55(51)3308.5942



Esta apostila é licenciada sobre a licença CC BY - Creative Commons Attribution 4.0 International License.

#### Dados internacionais de catalogação na publicação Lucia Helena Cunha Vidal CRB 10/2035

V285 Vanz, Samile Andréa de Souza  
Repositórios de dados de pesquisa / Pesquisadores Samile Andréa de Souza Vanz ; Caterina Groposo Pavão ; Rafael Port da Rocha ; Rene Faustino Gabriel Junior ; Sônia Elisa Caregnato; Bolsistas Ana Júlia Lopes ; Lucia Helena Cunha Vidal; Wagner Silva Wessfl. – Porto Alegre, RS : UFRGS, 2023.  
58 f.

Curso 2 da trilha **Ciência Aberta** abrange apostilas criadas pelo grupo RDPBrasil para treinamentos promovidos pela Rede Nacional de Pesquisa (RNP).

1. Repositório de dados de pesquisa. 2. Gestão de repositórios de dados de pesquisa. 3. Políticas de dados de pesquisa. Título. II. Pavão, Caterina Groposo. III. Rocha, Rafael Port da. IV. Gabriel Junior, Rene Faustino. V. Caregnato, Sônia Elisa. VI. Lopes, Ana Júlia. VII. Vidal, Lucia Helena Cunha. VIII. Wessfl, Wagner Silva.

#### **Como citar:**

VANZ, S. A. de S.; PAVÃO, C. G.; ROCHA, R. P. da; GABRIEL JUNIOR, R. F.; CAREGNATO, S. E.; LOPES, A. J.; VIDAL, L. H. C.; WESSFLL, W. S. Repositórios de dados de pesquisa. Porto Alegre: RDPBrasil, 2023.



ibict



## Curso 2

# Repositórios de dados de pesquisa

## MÓDULO I

### Objetivos

Apresentar o contexto em que os repositórios de dados de pesquisa estão inseridos, bem como citar a infraestrutura tecnológica que envolve essas ferramentas, trazendo como exemplos os *softwares DSpace e Dataverse*.

### Conceitos

Repositórios de Dados; *DSpace; Dataverse*.

### 1.1 REPOSITÓRIOS DE DADOS

A pesquisa científica contemporânea, ao produzir e utilizar intensamente dados de pesquisa, apresenta desafios e oportunidades. Os desafios consistem na inclusão da otimização do uso e reuso desses dados, e as oportunidades dizem respeito ao desenvolvimento de infraestruturas apropriadas ao seu armazenamento, gestão e preservação. Para que esses dados atendam às premissas de acesso, compartilhamento e reuso, eles requerem um ambiente digital que contemple uma infraestrutura tecnológica e de gestão ao longo do seu ciclo de vida.

O valor que os dados atribuem à pesquisa científica, de acordo com Borgman (2010), está relacionado à categoria a qual pertencem, sendo que alguns tipos de dados têm valor imediato e duradouro, outros ganham valor ao longo do tempo, e outros são mais fáceis de recriar do que selecionar. Diferentes dados são gerados nas pesquisas, e o seu compartilhamento impulsiona uma nova ciência, melhorando sua reprodutibilidade. No entanto, ao não se conseguir prever o valor futuro de um conjunto de dados, perde-se a noção de quanto o não

compartilhamento desses dados poderá vir a ser uma objeção para os cientistas futuramente.

Dessa forma, para apoiar a execução dos processos de gestão de dados, é necessária uma estrutura que compreenda todo o ciclo de vida dos dados. Neste contexto surgem os repositórios digitais de dados que, por muitas razões e demandas, se tornam parte essencial da infraestrutura mundial de pesquisa. Os repositórios de dados se destacam entre os sistemas informacionais com infraestrutura tecnológica e gerencial que apoiam os processos de gestão dos dados gerados nas atividades científicas, que são mantidos por ações que viabilizam o armazenamento de dados e incorporam uma infraestrutura necessária para a administração e preservação dos dados.

Um repositório de dados é uma infraestrutura responsável por armazenar conjuntos de dados de pesquisa, além de garantir o acesso aberto e a preservação aos resultados de pesquisa que surgem na forma de dados.

Sayão (2017) apresenta oito benefícios do armazenamento de dados de pesquisa em repositórios de dados, que são:

- a) ampliar a visibilidade dos dados de pesquisa;
- b) possibilitar o compartilhamento;
- c) dar crédito e reconhecimento aos autores dos dados;
- d) preservar os dados;
- e) contribuir para o princípio da transparência;
- f) colaborar com o armazenamento seguro;
- g) apoiar a curadoria digital;
- h) aumentar o reuso dos dados.

## EXEMPLOS DE REPOSITÓRIOS DE DADOS DE PESQUISA DO BRASIL

**Figura 1 - Repositório Arca Dados**

The screenshot shows the homepage of the ARCA dados FIOCRUZ repository. At the top, there is a navigation bar with the FIOCRUZ logo, the ARCA dados logo, and links for 'Página Inicial' and 'DataVerses'. Below the navigation bar, there are links for 'Pesquisa', 'Sobre', 'Guia do usuário', 'Suporte', 'Idioma Padrão', 'Inscreva-se', and 'Iniciar sessão'. The main content area features the ARCA dados FIOCRUZ logo and a description of the repository's purpose: to archive, publish, disseminate, preserve, and share digital research data. A search bar with the placeholder 'Digite sua pesquisa' and a 'Buscar' button is present. Below the search bar, there are links for 'Explorar' and 'Busca avançada'. Four statistics boxes are displayed: '71 Comunidades Dataverses', '726 Conjunto de dados', '194588 Arquivos', and '43022 Métricas / Downloads'. A section titled 'Assunto' lists various scientific fields with their respective counts: 'Medicina, Saúde e Ciências da Vida (767)', 'Ciências da Terra e do meio ambiente (682)', 'Ciências Sociais (74)', 'Artes e Humanidades (27)', and 'Computação e Ciência da Informação (10)'.

**Fonte:** FIOCRUZ, <https://arcadados.fiocruz.br/>

Figura 2 - Repositório Vitrine RNP Dados Abertos

Fonte: RNP, <https://dadosabertos.rnp.br/>

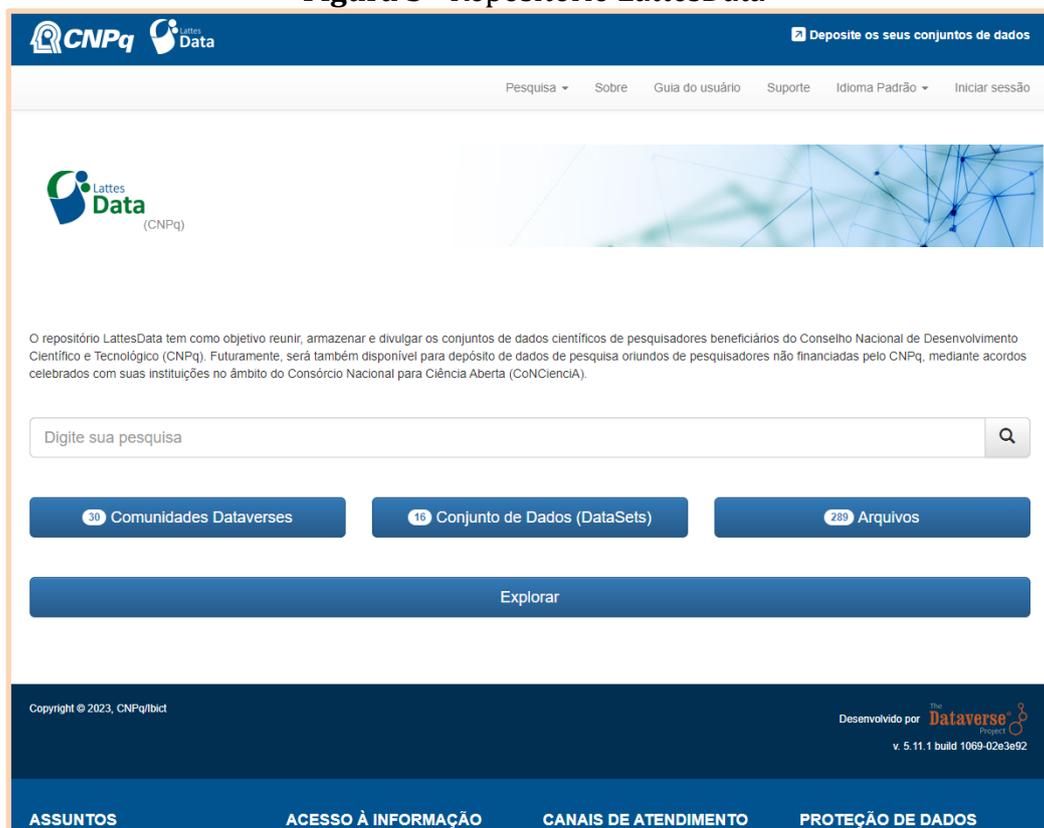
A disponibilização dos dados de pesquisas financiadas pelas agências de fomento no contexto brasileiro começa a ser discutida, apesar de ainda não ser mandatória em todas as agências. Têm-se iniciativas como a da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) que reconhece a importância da gestão adequada dos dados de pesquisa como parte essencial das boas práticas de pesquisa. Para tanto, considera necessário que os dados resultantes de projetos financiados pela Fundação sejam gerenciados e compartilhados de forma a garantir o maior benefício possível para o avanço científico, tecnológico, socioeconômico e cultural.

**Veja mais informações em:** <https://fapesp.br/14974/diretrizes-para-planos-de-gestao-de-dados-pgd-para-propostas-de-centros>

Outra agência de fomento à pesquisa que reconhece a importância da disponibilização e depósito dos dados de pesquisa é o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Neste sentido, a agência lançou o repositório de dados *LattesData* que busca reunir, armazenar e divulgar o conjunto de dados científicos de pesquisadores com projetos fomentados pelo CNPq.

Recentemente já é possível observar que a agência aderiu a ideia de disponibilização dos dados de pesquisa, como aconteceu na Chamada CNPq n.º 14/2023 – Apoio a Projetos Internacionais de Pesquisa Científica, Tecnológica e de Inovação: <https://www.gov.br/cnpq/pt-br/assuntos/noticias/cnpq-em-acao/submissao-aberta-para-chamada-de-apoio-a-projetos-intencionais-de-pesquisa-cientifica-tecnologica-e-de-inovacao>.

**Figura 3 - Repositório LattesData**



**Fonte:** CNPq, <https://lattesdata.cnpq.br/>

Figura 4 - Plano de Gestão de Dados da FAPESP

**FAPESP** Fomento à Pesquisa Pesquisa para Inovação Difusão do Conhecimento Boas Práticas e Políticas Sobre a FAPESP Converse com a FAPESP Índice Buscar...

**Gestão de Dados** Home > Financiamento à pesquisa > Gestão de Dados

**Índice**

- Gestão de Dados
- Conteúdo do Plano de Gestão de Dados - FAPESP
- Diretrizes para Planos de Gestão de Dados para propostas de Centros (CEPID, CPE, CPD)
- Ferramentas online para criação de Planos
- Documentos e páginas de interesse - Planos de Gestão de Dados
- Open Science @ FAPESP
- Webinário sobre como fazer um PGD
- Contato para dúvidas:** plano-gestao-dados@fapesp.br (ao enviar mensagem, favor indicar nome da sua instituição)

**Conteúdo do Plano de Gestão de Dados - FAPESP**

Para determinadas modalidades e chamadas, o documento "Plano de Gestão de Dados" faz parte dos anexos obrigatórios de uma proposta submetida à FAPESP.

A FAPESP não tem um modelo próprio para tais planos, pois estes variam conforme o domínio do conhecimento, os tipos de dados considerados e como os responsáveis pelo projeto pretendem disponibilizá-los. Algumas chamadas FAPESP poderão especificar o formato desejado do Plano.

As chamadas para Centros devem seguir as diretrizes específicas em "Diretrizes para Planos de Gestão de Dados para Propostas de Centros (CEPID, CPE, CPD)". Para todos os demais casos, o Plano submetido como anexo de uma proposta à FAPESP deve seguir as seguintes diretrizes:

Texto de até duas páginas, contendo as seguintes informações

- Descrição dos dados e metadados produzidos pelo projeto - por exemplo, amostras, registros de coleta, formulários, modelos, resultados experimentais, software, gráficos, mapas, vídeos, planilhas, gravações de áudio, bancos de dados, material didático e outros.
- Quando aplicável, restrições legais ou éticas para compartilhamento de tais dados, políticas para garantir a privacidade, confidencialidade, segurança, propriedade intelectual e outros.
- Política de preservação e compartilhamento (por exemplo, compartilhamento imediato ou apenas após a aceitação da publicação associada). Período de carência (antes do compartilhamento) e período durante o qual os dados serão preservados e disponibilizados.
- Descrição de mecanismos, formatos e padrões para armazenar tais itens de

**Gestão de Dados**

A FAPESP reconhece a importância da gestão adequada dos dados de pesquisa como parte essencial das boas práticas de pesquisa. Para tanto, considera necessário que os dados resultantes de projetos financiados pela Fundação sejam gerenciados e compartilhados de forma a garantir o maior benefício.

Fonte: FAPESP (2023).

## 1.2 TECNOLOGIAS PARA REPOSITÓRIOS DE DADOS

Repositórios de dados de pesquisa desempenham um papel fundamental na gestão e preservação de conjuntos de dados científicos, permitindo que pesquisadores, instituições e comunidades tenham acesso fácil e seguro a informações valiosas. Existem diversos *softwares* para a criação de repositórios de dados de pesquisa, entretanto, os *softwares* mais utilizados são o *DataVerse* e o *DSpace*.



O *software* **Dataverse** é uma plataforma de código aberto projetada para permitir o compartilhamento, publicação e preservação de conjuntos de dados de

pesquisa. Ele fornece uma infraestrutura para pesquisadores e instituições armazenarem, gerenciarem e compartilharem dados de pesquisa de maneira organizada e de acordo com as melhores práticas.

O *Institute for Quantitative Social Science* (IQSS) colabora com a *Harvard University Library* e a organização *Harvard University Information Technology* para o desenvolvimento do *software* Dataverse. Esse *software* permite a colaboração entre desenvolvedores de todo o mundo através da governança do Consórcio Global da Comunidade do *Dataverse* – GDCC, em inglês, *Global Dataverse Community Consortium* (<https://dataversecommunity.global/>). O *Dataverse* foi criado com base na experiência do projeto anterior chamado *Virtual Data Center* (VDC), que foi executado de 1999 a 2006 como uma colaboração entre o *Data Center Harvard-MIT* e a Biblioteca da Universidade de Harvard.

O *Dataverse* é um aplicativo desenvolvido em *Java Enterprise Edition* (EE) para a *web*, fornecido como um arquivo WAR (*Web Archive*). Ele é predominantemente escrito em Java e utiliza o servidor de aplicação *Glassfish* como seu serviço de *back-end*.

O *Dataverse* permite aos usuários criar repositórios de dados onde os conjuntos de dados podem ser carregados, documentados, explorados e disponibilizados para acesso público ou restrito. Ele oferece recursos para metadados detalhados, versionamento de dados, controle de acesso, geração de identificadores persistentes de objetos digitais como DOIs (*Digital Object Identifiers*), integração com outros sistemas de armazenamento e muito mais.

Assim, um repositório de dados *Dataverse* consiste em uma instalação de *software* capaz de abrigar múltiplas comunidades de dados, também conhecidas como *Dataverses* no *software Dataverse*. Ou seja, no *software Dataverse*, cada comunidade *Dataverse* representa uma comunidade de dados, a qual contém conjuntos de dados, chamados *Datasets*. Além dos conjuntos de dados tem-se a informação de metadados descritivos correspondente, documentações associadas e informação do versionamento dos arquivos. Além disso, uma comunidade *Dataverse* pode conter várias comunidades filhas com conjunto de dados específicos.

Atualmente, A GDCC conta com aproximadamente 100 instituições usuárias, com instalações de *Dataverses* distribuídas nas Américas do Norte, Central e do Sul, Europa, África, Ásia e Oceania. A iniciativa mantém a *Dataverse Users Community* (<https://groups.google.com/g/dataverse-community>) para troca de experiências e informações.

Um dos objetivos principais do *software Dataverse* é facilitar a descoberta e o acesso a dados de pesquisa, promovendo a transparência, a colaboração e a reprodutibilidade científica. Ele é usado por pesquisadores, instituições acadêmicas, organizações governamentais e outras entidades interessadas em gerenciar e compartilhar seus dados de pesquisa de maneira eficiente e confiável.

**DSPACE**

O **DSpace** é um *software* de código aberto e uma plataforma de repositório digital

amplamente utilizado para gerenciar, preservar e compartilhar conteúdos digitais, como documentos, artigos, teses, imagens, vídeos, conjuntos de dados e outros tipos de materiais. Ele foi desenvolvido inicialmente pela Universidade de Massachusetts e agora é mantido e aprimorado por uma comunidade de desenvolvedores e instituições em todo o mundo.

O objetivo principal do *DSpace* é fornecer uma solução robusta e flexível para criar repositórios digitais institucionais, bibliotecas digitais e arquivos abertos. Ele oferece uma plataforma confiável para armazenar, organizar e disseminar conteúdo digital, tornando-o acessível a pesquisadores, estudantes e ao público em geral.

O *DSpace* possui interface de usuário amigável e suporta metadados detalhados para descrever adequadamente os itens armazenados, facilitando a descoberta e a indexação em mecanismos de busca. O *software* oferece opções para preservar e arquivar digitalmente os itens armazenados, garantindo sua acessibilidade e durabilidade a longo prazo.

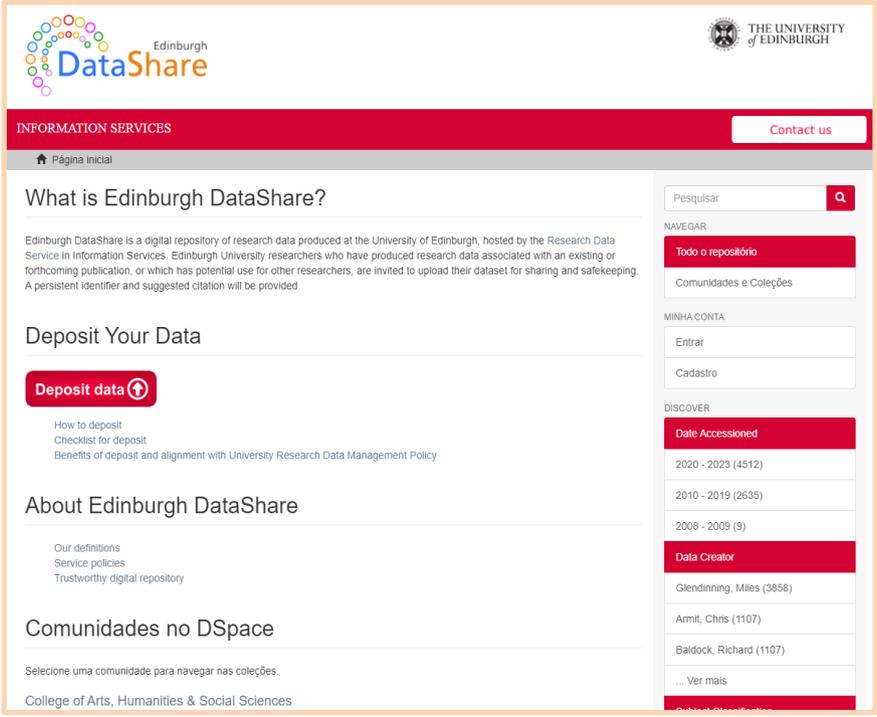
No Brasil, o *software* foi amplamente divulgado pelo Ict que coordenou um movimento entre as bibliotecas universitárias para a criação de repositórios institucionais de publicações científicas nos anos 2000.

## REPOSITÓRIOS BASEADOS EM *DSPACE*

[Base de Dados Científicos](#)  
**Universidade Federal do Paraná**



## REPOSITÓRIOS BASEADOS EM DSPACE



[Edinburgh Data Share](https://www.ed.ac.uk/data-share)  
**University of Edinburgh**





### Recomendação de leitura:

GABRIEL JUNIOR, R. F.; ROCHA, R. P. da; CAREGNATO, S. E.; PAVÃO, C. M. G.; PASSOS, P. C. S. J.; BORGES, E. N.; VANZ, S. A. de S.; AZAMBUJA, L. A. B. Acesso aberto a dados de pesquisa no Brasil: mapeamento de repositórios, práticas e percepções dos pesquisadores e tecnologias. **Ciência da Informação**, [S. l.], v. 48, n. 3, 2020. DOI: 10.18225/ci.inf.v48i3.4958. Disponível em: <https://revista.ibict.br/ciinf/article/view/4958>. Acesso em: 13 set. 2023.

**Figura 5** – Acesso aberto a dados de pesquisa no Brasil: mapeamento de repositórios, práticas e percepções dos pesquisadores e tecnologias

## **Acesso aberto a dados de pesquisa no Brasil: mapeamento de repositórios, práticas e percepções dos pesquisadores e tecnologias**

**Rene Faustino Gabriel Junior**  
Doutor em Ciência da Informação pela Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp) – Marília, SP – Brasil. Professor da Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, RS - Brasil.  
<http://lattes.cnpq.br/5900345665779424>  
<http://orcid.org/0000-0003-1021-3360>  
E-mail: [renefgj@gmail.com](mailto:renefgj@gmail.com)

**Rafael Port da Rocha**  
Doutor em Ciência da Computação pela Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, RS - Brasil. Professor da Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, RS - Brasil.  
<http://lattes.cnpq.br/5118387541734094>  
<http://orcid.org/0000-0003-4131-0534>  
E-mail: [rafael.rocha@ufrgs.br](mailto:rafael.rocha@ufrgs.br)

**Sônia Elisa Caregnato**  
Doutora em Information Studies pela University of Sheffield (SHEFFIELD) - Inglaterra. Professora da Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, RS - Brasil.  
<http://lattes.cnpq.br/5627209208288722>  
<https://orcid.org/0000-0002-5676-2763>  
E-mail: [sonia.caregnato@ufrgs.br](mailto:sonia.caregnato@ufrgs.br)

**Caterina Marta Groposo Pavão**  
Doutora em Comunicação e Informação pela Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, RS – Brasil, com período sanduíche em Universidad Complutense de Madrid (UCM) - Espanha. Professora da Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, RS - Brasil.  
<http://lattes.cnpq.br/4834791532698069>  
<http://orcid.org/0000-0003-3712-7200>  
E-mail: [caterina@cpd.ufrgs.br](mailto:caterina@cpd.ufrgs.br)

**Paula Caroline Schifino Jardim Passos**  
Pós-Doutorado pela Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, RS - Brasil. Doutora em Comunicação e Informação pela Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, RS - Brasil.  
<http://lattes.cnpq.br/1766582596876949>  
<http://orcid.org/0000-0002-2354-847X>  
E-mail: [paulacarolinejardim@gmail.com](mailto:paulacarolinejardim@gmail.com)

**Eduardo Nunes Borges**  
Pós-Doutorado pela Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, RS - Brasil. Doutor em Computação pela Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, RS - Brasil. Professor da Universidade Federal do Rio Grande (FURG) - Rio Grande, RS – Brasil.

**Fonte:** RDP Brasil (2019). Disponível em:  
<https://lume.ufrgs.br/handle/10183/212266>

**Figura 6** – Repositórios de dados de pesquisa



**Fonte:** RDP Brasil / IBICT (2021).

**Para ir além:** *Explore os softwares*

*Dataverse:* <http://dataverse.org/>

*DSpace:* <http://www.dspace.org/>

## MÓDULO II

### Objetivos

Compreender a arquitetura dos repositórios de dados de pesquisa, tanto no *DSpace* quanto no *Dataverse*.

### Conceitos

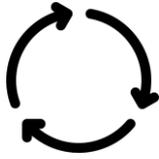
Repositório de Dados de Pesquisa; Arquitetura da Informação; *DSpace*; *Dataverse*.

## 2.1 ARQUITETURA DO REPOSITÓRIO E ORGANIZAÇÃO DAS COMUNIDADES

A arquitetura de dados desempenha um papel fundamental no planejamento e na implementação de uma infraestrutura de dados eficiente, considerando todas as etapas do ciclo de vida dos dados. Ela aborda de forma abrangente e integrada a necessidade de estabelecer padrões de metadados, estruturas sintáticas e semânticas, bem como garantir a conformidade com princípios éticos, legais e técnicos. Essas diretrizes são essenciais para permitir o consumo e o fornecimento de dados para diversas finalidades, fazendo uso de técnicas e tecnologias inovadoras e emergentes. Através da arquitetura de dados, é possível criar um ambiente sólido e coeso que garante a qualidade, a segurança e a interoperabilidade dos dados, possibilitando seu efetivo uso e compartilhamento (TORINO, 2023).

A arquitetura de um repositório de dados de pesquisa refere-se à estrutura, aos componentes e às tecnologias utilizadas para gerenciar, armazenar e disponibilizar os dados de pesquisa (DATA..., 2023). Ela engloba os aspectos técnicos e organizacionais envolvidos na criação e manutenção do repositório. A

arquitetura de um repositório de dados de pesquisa deve considerar os seguintes elementos:



**CICLO DE VIDA DOS DADOS DE PESQUISA:** “[...] compreender como ocorrem as fases de planejamento, coleta e disponibilização dos dados, que envolvem as formas com que são criados ou coletados, tratados, armazenados, disponibilizados para consulta e, quando necessário, descartados, etapas essas que compõem o ciclo de vida dos dados.” (TORINO, 2023, p. 193-194).



**METADADOS:** “[...] metadados são utilizados para representar e organizar a informação contida em um objeto, visando melhorar o acesso [...]” (TORINO, 2023, p. 209).



**ESTRUTURA SINTÁTICA E SEMÂNTICA DOS DADOS:** “É por meio da semântica dos dados que os dados armazenados em diferentes fontes serão compatibilizados para o modelo de dados utilizado, o que requer que haja documentação acionável por humanos e máquinas visando estabelecer a compreensão.” (TORINO, 2023, p. 231).

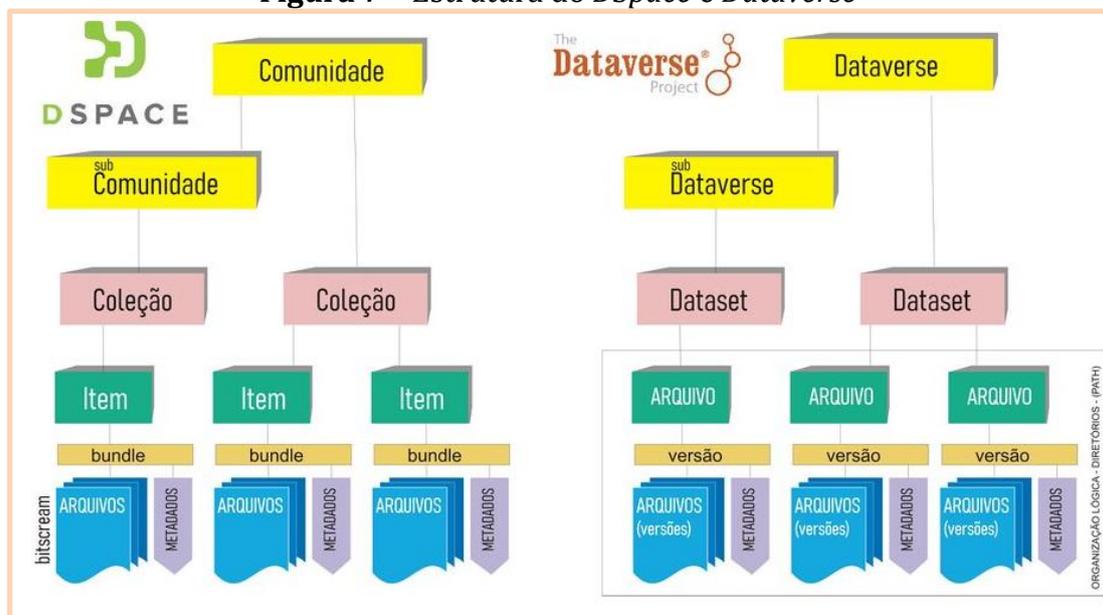


**PRINCÍPIOS ÉTICOS, LEGAIS E TÉCNICOS APLICADOS AOS DADOS DE PESQUISA:** diz respeito às leis, regulações e segurança, “[...] estão fortemente atrelados à coleta, ao tratamento, ao armazenamento e à disponibilização de dados, considerando os diferentes contextos, tecnologias e finalidades.” (TORINO, 2023, p. 251).

## 2.2 ARQUITETURA DOS REPOSITÓRIOS EM DATAVERSE E DSPACE

A estrutura informacional do *DSpace* e do *Dataverse* é hierárquica, composta por comunidades, coleções e itens. Essa estrutura permite a organização de acervo, e também, facilita a recuperação dos objetos digitais depositados, considerando que muitos usuários navegam pelo repositório a partir da estrutura das comunidades. Apesar de ambos os *softwares* possibilitarem diversas opções e filtros para pesquisa, considera-se importante a organização hierárquica e visual do repositório, tendo em vista que a busca por navegação é uma opção de diversos usuários.

**Figura 7 – Estrutura do DSpace e Dataverse**



**Fonte:** BRAPCI (2021).

As comunidades *Dataverse* e subcomunidades *Dataverse* são estruturas informacionais que representam a organização do repositório. As comunidades são as estruturas de mais alto nível e podem conter vários níveis de subcomunidades. Assim, representam apenas a estrutura, não contendo objetos digitais diretamente. Na tradução do *software Dataverse* para a língua

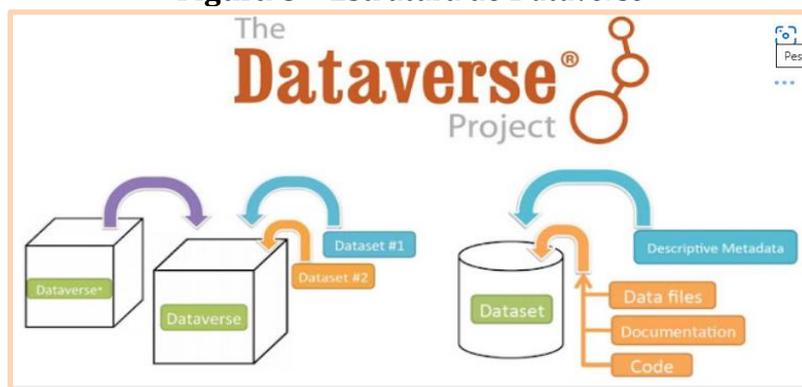
portuguesa, optou-se pelo uso do termo comunidades e subcomunidades, para facilitar o entendimento dos usuários.

Nesse contexto, as comunidades e subcomunidades podem representar temas ou estruturas organizacionais. Por exemplo, em um repositório institucional de uma universidade pode ser organizado pela estrutura organizacional da instituição. Neste caso, as comunidades podem representar as faculdades e institutos, enquanto as subcomunidades representam os departamentos ou programas de pós-graduação. Outra opção é organizar o repositório por área do conhecimento e neste caso, as comunidades poderiam representar as grandes áreas, enquanto as subcomunidades representariam as subáreas. É possível criar quantas comunidades forem necessárias para representar o acervo, e em cada comunidade pode-se criar tantas subcomunidades, em vários níveis, quantas forem necessárias para refinar essa comunidade.

No caso do *DSpace*, as coleções são agrupamentos de documentos, são estruturas que servem, preferencialmente, para agrupar documentos com alguma característica comum. Toda coleção deve pertencer a uma comunidade ou subcomunidade, pois enquanto as comunidades organizam o repositório, as coleções organizam os documentos do acervo.

O equivalente às coleções no *Dataverse* são os *Datasets* ou Conjunto de Dados. *Dataset* é um contêiner de muitos arquivos de dados, documentação e códigos. Para cada *dataset* é atribuído um identificador de objeto digital (DOI) exclusivo. A unidade informacional é chamada de item, no caso do *DSpace*, e arquivo de dados, no caso do *Dataverse*.

**Figura 8 – Estrutura do Dataverse**

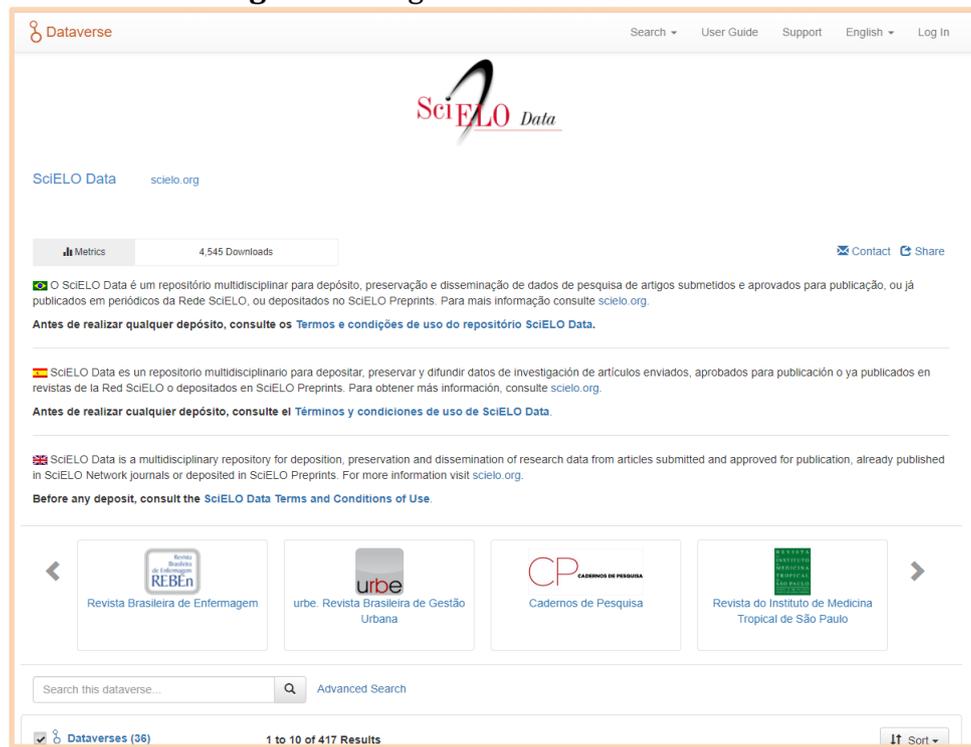


Fonte: BRAPCI (2021).

**EXEMPLOS:**

O **SciELO Data** é um repositório multidisciplinar para depósito, preservação e disseminação de dados de pesquisa de artigos submetidos e aprovados para publicação, ou já publicados em periódicos da Rede SciELO, ou depositados no SciELO Preprints. A organização acontece a partir das revistas, e dentro delas, as comunidades são organizadas a partir dos artigos publicados pelas revistas.

**Figura 9 – Página inicial do SciELO Data**



Fonte: SciELO (2023).

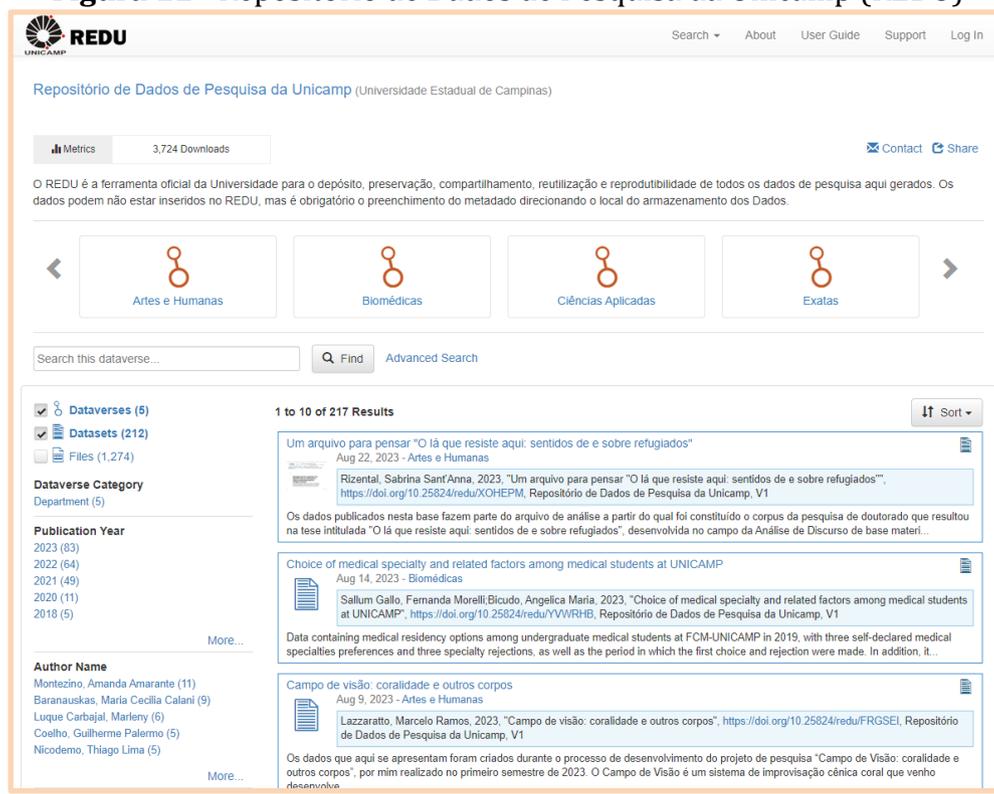
**Figura 10** – Revista Brasileira de Enfermagem indexada no *SciELO Data*

The screenshot displays the SciELO Data interface for the journal 'Revista Brasileira de Enfermagem'. At the top, there is a navigation bar with 'Dataverse' and utility links like 'Search', 'User Guide', 'Support', 'English', and 'Log In'. The journal's logo and name are prominently featured, along with its website 'scielo.br/reben'. Below this, there are instructions in multiple languages (Brazilian Portuguese, Spanish, and English) regarding data deposition and the use of SciELO Data terms. A search bar is present with a search button and an 'Advanced Search' link. On the left, there are filters for 'Dataverses (0)', 'Datasets (74)', and 'Files (254)', along with a 'Publication Year' filter showing 32 results for 2023 and 42 for 2022. The main content area shows '1 to 10 of 74 Results'. The first result is a dataset for 'REBEn 2022-0721: Competências de Liderança do Enfermeiro Especialista em Enfermagem Médico-Cirúrgica' dated Aug 23, 2023. The dataset description includes the authors (Pires, Marina; Lopes, Romana Silva; Caetano, Cláudia Sofia Fernandes; Mota, Lilliana; Príncipe, Fernanda) and a brief abstract: 'This is a quantitative, descriptive study involving 311 Nurse Specialists in Medical-Surgical Nursing, in a non-probabilistic convenience sample. Inclusion criteria were having the title of Specialist Nurse in Medical-Surgical Nursing as well as exercising the profession in Portu...'. A second result is partially visible at the bottom: 'REBEn-2022-0767: Cuidado de pessoas idosas com incapacidades em Instituições de Longa Permanência para Idosos'.

**Fonte:** SciELO (2023).

O **Repositório de dados de pesquisa da Unicamp (REDU)** é um exemplo de repositório organizado por grandes áreas do conhecimento. Neste caso, são cinco comunidades (*dataverses*), e dentro delas, a organização acontece por projeto de pesquisa.

**Figura 11 - Repositório de Dados de Pesquisa da Unicamp (REDU)**



Fonte: Unicamp (2023).

O **Repositório Arca Dados**, da Fiocruz, está organizado por unidade administrativa da Fiocruz.

**Figura 12 - Repositório Arca Dados organizado por unidade administrativa da Fiocruz**



Fonte: Fiocruz (2023).

## MÓDULO III

### Objetivos

Visa ressaltar a importância da construção de políticas para a gestão de dados nos repositórios de dados de pesquisa.

### Conceitos

Repositório de Dados de Pesquisa; Gestão de Dados; Políticas.

### 3.1 INSTRUMENTOS DE PLANEJAMENTO E GESTÃO DE REPOSITÓRIOS DE DADOS DE PESQUISA

A gestão de repositório de dados de pesquisa refere-se ao conjunto de atividades e processos envolvidos na organização, armazenamento, preservação e compartilhamento de dados de pesquisa. Dessa forma, a gestão de um repositório de dados de pesquisa objetiva facilitar o acesso, a reutilização e a transparência dos dados, promovendo a colaboração e o avanço da pesquisa científica. Diversos instrumentos são utilizados para gestão de repositórios, entre eles, podemos mencionar:



**POLÍTICAS:** “[...] ou diretrizes são planos gerais de ação, guias genéricos que definem linhas mestras, orientam a tomada de decisão e dão estabilidade à organização.” (ALMEIDA, 2005, p. 6).



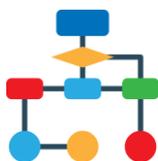
**MANUAL:** Um manual é um documento ou guia que fornece instruções, orientações ou informações sobre como realizar uma tarefa, operar um dispositivo, usar um produto ou seguir determinados procedimentos. Geralmente, um manual é organizado

de forma estruturada e clara, apresentando passo a passo as etapas necessárias para a realização de uma atividade específica. (UNIVERSIDADE..., 1999).



**MANUAL DE PROCEDIMENTOS:** “[...] são instrumentos que estabelecem métodos rotineiros de execução de atividades e detalham a maneira exata pela qual uma atividade deve ser realizada e a sequência em que essas rotinas são realizadas.”

(ALMEIDA, 2005, p. 6).



**FLUXOGRAMA:** “Diagrama esquemático de um processo que ilustra, de forma descomplicada, a transição de informações entre os elementos que o compõem. Na prática, o fluxograma é feito a partir do mapeamento das atividades e descreve o passo a passo necessário para a execução de um processo.” (BATISTA, 2016, p. 28).



**REGULAMENTO:** É um conjunto de regras, diretrizes ou normas estabelecidas por uma autoridade ou organização para governar o comportamento, a conduta ou o funcionamento de um determinado contexto, como uma instituição, setor ou atividade específica

(UNIVERSIDADE..., 2023).



**ORGANOGRAMA:** “Gráfico que reflete a estrutura hierárquica de uma organização social complexa, representa simultaneamente os diferentes elementos dos grupos e suas ligações.” (BATISTA, 2016, p. 28). Assim, organograma é uma representação gráfica da

estrutura organizacional de uma empresa, instituição ou organização. Mostra as relações hierárquicas entre os diferentes cargos, departamentos e unidades dentro da organização. Os organogramas são comumente usados para facilitar a compreensão da estrutura organizacional, mostrar as relações de autoridade e responsabilidade, e auxiliar na comunicação interna e na tomada de decisões.



**CRONOGRAMA:** “Disposição gráfica do tempo que será gasto na realização de um trabalho, de acordo com as atividades a serem cumpridas. O cronograma serve para auxiliar no gerenciamento e controle, permitindo, de forma rápida, a visualização de seu andamento.” (BATISTA, 2016, p. 28).

Neste curso serão apresentadas e discutidas algumas sugestões para a redação dos principais instrumentos para gestão de um repositório de dados de pesquisa: a política geral, os termos de uso do repositório (*download* e *upload*) e um fluxograma de trabalho. Não há uma regra que estabeleça instrumentos obrigatórios, mas percebe-se que quanto maior o repositório e a comunidade por ele atendida, maior o volume de regras necessárias para seu bom funcionamento. Ao analisar repositórios internacionais, observa-se que existe no mínimo uma política geral, no entanto, vários repositórios lançam mão de instrumentos adicionais (SILVEIRA; VANZ, 2023).

O objetivo de um repositório de dados de pesquisa é arquivar e publicar dados de pesquisas produzidas pela comunidade científica e/ou produzidos pela comunidade em parceria com terceiros. Basicamente, consideram-se três tipos de usuários. É fundamental pensar no funcionamento do repositório a partir do ponto de vista de cada uma dessas três categorias de usuários, e quais regras cada um desses usuários vai seguir:



a primeira categoria de usuário é mais simples, é aquele que vai acessar o repositório, baixar os dados para reuso;



o segundo tipo de usuário é mais complexo, que é o pesquisador que vai depositar dados no repositório;



o terceiro tipo de usuário é o chamado curador de dados. O curador tem a responsabilidade de avaliar cada depósito de dados em relação aos metadados descritos, e também acerca dos aspectos legais e éticos, para que a publicação e o compartilhamento aconteçam de forma segura.

Para garantir que todas as regras sejam seguidas, e que sejam consideradas todas as questões éticas e legais, é necessário ter um conjunto de documentos essenciais que estabeleçam as condições básicas de funcionamento do repositório. Esses documentos devem mencionar os instrumentos legais relevantes para o contexto da instituição que hospeda e gerencia o repositório. A existência de políticas em repositórios publiciza aos usuários as normas adotadas pelos repositórios, proporcionando assim segurança e confiabilidade em relação ao depósito de seus dados de pesquisa e ao acesso e uso destes por terceiros (SILVEIRA; VANZ, 2023). A falta de adequação e clareza dessas políticas podem ocasionar uma subutilização dos repositórios, por insegurança dos usuários (CARVALHO e CARVALHO, 2012).

### 3.2 POLÍTICAS PARA REPOSITÓRIOS DE DADOS DE PESQUISA

As políticas dos repositórios são um conjunto de diretrizes que devem ser seguidas por aqueles que acessam e utilizam os serviços oferecidos. Esses documentos também fornecem orientações sobre as práticas utilizadas pelo repositório para armazenar e preservar seus conteúdos, além de abordar questões técnicas, legais e de infraestrutura relacionadas à gestão: “As políticas devem refletir e estar em consonância com os princípios e diretrizes organizacionais da instituição à qual o repositório está vinculado, uma vez que ele é uma extensão da organização.” (SILVEIRA, 2021, p. 129).

Os elementos a serem considerados nas políticas de repositórios são bastante abrangentes, relacionados à própria instituição, como por exemplo, as disciplinas

abrangidas e a qual órgão da universidade cabe a gerência e diretrizes institucionais do repositório; assim como elementos relacionados aos dados em si, como por exemplo, informações acerca da proveniência dos dados, se os dados são oriundos de pesquisadores com vínculo institucional ou de pesquisadores externos ou afiliados, se há restrição de depósito, entre outras informações.

O contexto institucional deve ser considerado ao redigir as políticas de um repositório de dados de pesquisa. Em geral, o documento inicia com uma apresentação do repositório, sua missão, valores e objetivos. A seguir apresenta-se uma breve orientação para a redação de uma política de repositório de dados. São sugeridas algumas seções que podem compor o documento da política, mas não há regras que contemplem elementos mínimos ou obrigatórios. A política é um documento que deve ser redigido dentro do contexto de cada instituição.



## ORIENTAÇÃO PARA REDAÇÃO DA POLÍTICA DE REPOSITÓRIO DE DADOS

### 1. Apresentação ou Introdução

Sugere-se redigir alguns parágrafos para introduzir informações gerais sobre o repositório e o contexto institucional. Por exemplo:

O Repositório [nome] é um sistema para dados de pesquisa da [... por exemplo, Universidade ou Instituição x ]. A visão do repositório é [...]. O objetivo do repositório é [... por exemplo, fornecer acesso aberto a dados resultantes das pesquisas realizadas por membros da comunidade nas áreas de atendidas pela instituição e dar visibilidade à produção, garantir a reprodutibilidade das pesquisas, preservar a produção, etc].

O Repositório de Dados de Pesquisa destina-se aos [...descrever os grupos de usuários, como por exemplo, pesquisador, alunos, professores, comunidade em geral...].

O repositório utiliza a plataforma [...por exemplo, Dataverse, desenvolvida e mantida pelo Dataverse Project, na Universidade de Harvard [link].

Esta política está em consonância com [se houver uma política institucional de acesso aberto/ciência aberta ou para gestão de dados, mencionar aqui, bem como outras legislações relevantes para a instituição].

## 2. Organização do repositório

Sugere-se redigir alguns parágrafos para descrever a arquitetura do repositório, como está organizado em comunidades e subcomunidades. É interessante incluir imagens e o organograma do repositório, além de indicar a responsabilidade pelas comunidades, especialmente nos casos em que várias pessoas ou equipes dividem responsabilidades. Destaca-se que o registro da criação das comunidades é importante para justificar o desenvolvimento do repositório, permitindo ou evitando a criação de comunidades de forma aleatória no futuro. O parágrafo a seguir pode orientar a redação do trecho:

O repositório [nome] é estruturado [citar como está estruturado, por exemplo, em comunidades] para a organização da informação e da produção científica de [por exemplo, da Universidade X, Instituição Y, ou de cada unidade do...]. Cada comunidade representa/corresponde à uma [unidade, departamento, grupo de pesquisa... da instituição]. A criação de novas comunidades é permitida quando [ou não é permitida]. Os responsáveis pela comunidade são [...].

## 3. Dados de pesquisa

Sugere-se redigir um texto para informar o tipo de dado que pode ser armazenado no repositório, por exemplo, qualquer tipo de dados; ou especificamente, dados estatísticos, imagens, texto simples, texto estruturado, dados audiovisuais, dados bibliográficos, som, entre outros tipos.

Sugere-se que esta seção informe os formatos de arquivos aceitos, por exemplo, .csv, .txt, .sav; as licenças utilizadas pelo repositório, e o modelo de disponibilização dos dados, se total, imediata, embargada, entre outros modelos.

As responsabilidades sobre o processo de arquivamento devem ser esclarecidas, evitando sobrecarga de tarefas em um único setor: se autoarquivamento, se é responsabilidade de um setor ou gestor de comunidade.

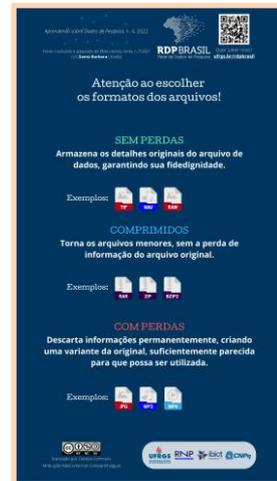
O repositório deve indicar necessidade de anonimização dos dados e controle / responsabilidade acerca de quaisquer informações sigilosas e sensíveis.

Recomenda-se a aplicação de licenças abertas, como as licenças *Creative Commons CC BY*, *Open Data Commons*, *AusGoal*, ou licenças específicas do próprio repositório. No entanto, a indicação do uso das *Creative Commons* é o mais comum e isso se deve ao fato de estas criarem instrumentos jurídicos flexíveis, de fácil compreensão de seu conteúdo, que se fundamentam nas prerrogativas que cada indivíduo tem como autor, de permitir o acesso e uso de suas obras (Lemos, 2005; Swan, 2016). A escolha da licença pode ficar a critério do autor dos dados, e também depende dos requisitos do financiador. Em alguns casos, a opção *CC0 Public Domain Dedication* (sem direitos reservados) pode ser mais apropriada.

O texto a seguir pode dar ideias para a redação:

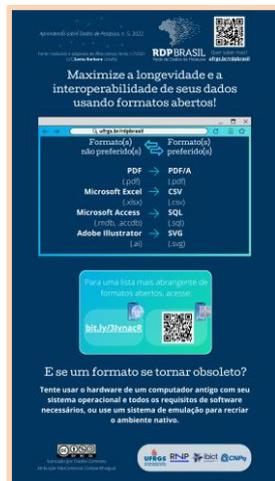
O repositório permite o depósito de quaisquer tipos de dados de pesquisa, como por exemplo [citar exemplos, como dados estatísticos, entrevistas, dados observacionais] em formatos [informar se permite formatos proprietários ou somente formatos abertos]. O depósito de dados resulta em [sua publicação imediata, ou embargo permitido pelo tempo x, ou outra possibilidade]. Para submissão os dados devem ser [se for o caso, anonimizados ou submetidos a algum tipo de processo específico]. Os dados estarão disponíveis sob licença [informar a licença adotada pelo repositório, ou informar se a decisão acerca de licença fica a cargo do depositante].

SAIBA MAIS:



**Porque os formatos abertos são importantes?**

**Atenção ao escolher os formatos dos arquivos!**



**Maximize a longevidade e a interoperabilidade de seus dados usando formatos abertos!**

**A importância da privacidade de dados pessoais**

**4. Metadados**

Nesta seção é preciso informar quem é responsável pela descrição dos metadados, se biblioteca, pesquisador, coordenador da comunidade, entre outras

possibilidades, além dos metadados mínimos para descrição dos conjuntos de dados (conforme nível de especificidade de cada repositório). Também é importante informar a licença que se aplica aos metadados. Veja o exemplo:

Os dados depositados devem ser descritos pelos metadados [autor, ano de criação, etc, informar todos os metadados obrigatórios] para que sejam identificados e permitam reuso. O processo de descrição é de responsabilidade de [se biblioteca, pesquisador, coordenador da comunidade, entre outras possibilidades]. Os metadados dos itens depositados no Repositório [nome] são disponibilizados sob a licença [por exemplo, *Creative Commons Attribution CC BY (Attribution Only)*].

## 5. Usuários e responsabilidades

Sugere-se descrever o público usuário do repositório de dados de pesquisa e responsabilidades. As responsabilidades podem ser descritas junto com o tipo de usuário ou em seção separada, indicando os usuários responsáveis pelas atividades como *upload* de arquivo, descrição de dados, revisão de dados, publicação de dados, atribuição de DOI, entre outras. Sugere-se indicar se os usuários devem se identificar para uso do repositório e como é feita essa identificação. Por exemplo:

O repositório [nome] considera os seguintes tipos de usuários [informar as categorias].

Usuários internos: [docentes, discentes, pesquisadores]. Podem depositar seus resultados de pesquisa no repositório, navegar e pesquisar os metadados e, na maioria dos casos, baixar o conteúdo de todas as coleções do repositório.

Usuários externos: [pesquisadores locais e internacionais, cidadãos, professores, estudantes, público em geral, etc]. Podem navegar e pesquisar os metadados e baixar os dados em acesso aberto do repositório.

Associados: [pesquisadores profissionais locais e internacionais que têm projetos conjuntos com usuários internos]. Podem navegar e pesquisar os metadados,

baixar o conteúdo do em acesso aberto do repositório, acessar a coleção dedicada ao seu projeto.

A identificação dos usuários é feita por [por exemplo, ORCID, rede CAFe, login da instituição].

## 6. Submissão

Sugere-se descrever quem é responsável pela submissão, e uma breve descrição do processo. É importante mencionar a possibilidade de auto depósito. Veja o exemplo:

Os usuários autorizados a fazer *upload* de dados incluem [bibliotecários, pessoal, pesquisadores, professores etc]. O depositante é responsável pela descrição de metadados [ou a biblioteca ou outro órgão o faz]. Metadados e itens a serem submetidos estão sujeitos a verificação pelos administradores do repositório, que examinam a elegibilidade de autores/depositantes, relevância para o escopo do Repositório, layout e formatos válidos.

Os depositantes devem possuir todos os direitos necessários para depositar os dados. Os depositantes de dados são responsáveis pelo cumprimento de quaisquer regulamentos legais nacionais ou internacionais e embargos de editores.

A validade e autenticidade do conteúdo das submissões e quaisquer violações de direitos autorais são de inteira responsabilidade do depositante.

Se o Repositório de Dados receber prova de violação de direitos autorais, o item relevante será removido o mais rápido possível.

Especificação de limite de tamanho dos arquivos, por exemplo: Arquivos com tamanho superior a [50] MB não podem ser carregados no repositório, exceto como permissão especial dos administradores do repositório.

Ao depositar os metadados e o conteúdo no Repositório de Dados, os depositantes aceitam os termos e condições definidos nesta Política do Repositório.

## 7. Versionamentos

Informar se o Repositório opera com possibilidade de versionamento de arquivos, as chamadas versões efêmeras ou em andamento; ou aceita somente o depósito de arquivo final e consolidado, por exemplo:

O Repositório [nome] aceita depósito de [versão consolidada dos dados de pesquisa; ou, se necessário, uma versão atualizada poderá ser depositada em substituição à versão anterior, que será retirada ao público; ou, são permitidos quantos versionamentos o autor julgar necessário].

## 8. Privacidade

Tendo em vista a LGPD, informar se o Repositório retém informações de acesso. Sugere-se que esta seção da política seja construída em conjunto com o setor de TI ou procuradoria da instituição. Algumas ideias para registrar nesta seção da política:

Sempre que um usuário visita o repositório, as seguintes informações são coletadas e armazenadas automaticamente [endereço de IP, tipo de navegador e sistema operacional, data e hora, páginas visitadas, localização, entre outros].

Essas informações são usadas internamente apenas para [solução de problemas técnicos, para melhorar a usabilidade do repositório e para registrar estatísticas agregadas]. O sistema coleta informações pessoais enviadas durante o processo de depósito. As informações enviadas serão usadas apenas para [o propósito para o qual você as enviou e não serão compartilhadas com terceiros ou usadas para fins comerciais].

O Google Analytics é usado para analisar o uso deste repositório. A política de privacidade do Google está disponível em: <http://www.google.com/privacypolicy.html>.

[Se você usa aplicativos de terceiros, você deve listá-los na política de privacidade e fornecer links para suas políticas relacionadas a dados pessoais.]

## 9. Preservação digital

Sugere-se informar o detalhamento do plano de preservação digital. Repositórios que possuem política específica para preservação podem indicar o *link*.

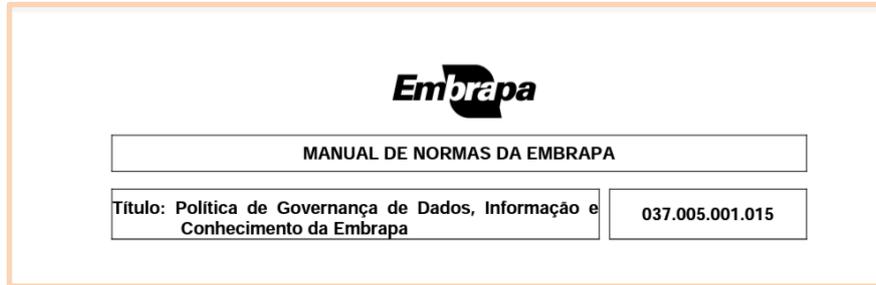
## 10. Responsáveis e data de aprovação da política

Sugere-se finalizar a política com a publicização da data de aprovação do documento e nome dos responsáveis pela redação, de forma a compartilhar responsabilidades. Sugere-se que representantes de diferentes setores da instituição façam parte do comitê gestor do repositório, por exemplo, biblioteca, setor de TI, procuradores ou advogados, docentes e alunos.

### EXEMPLOS DE POLÍTICA:



**Política de gestão, compartilhamento e abertura de dados para pesquisa:  
princípios e diretrizes (Fiocruz)**



**Política de Governança de Dados, Informação e Conhecimento  
(Embrapa)**

## MÓDULO IV

### Objetivos

Discute as políticas relacionadas ao *upload* (depósito).

### Conceitos

Política de *Upload*; Política de Depósito; Termos de Submissão.

### 4.1 POLÍTICA DE UPLOAD OU DEPÓSITO

Em muitos repositórios, a política geral contempla os detalhes de *upload* ou depósito. Entretanto, há repositórios que optam por redigir um documento específico, a política de *upload* ou depósito, que reúne as regras para depósito dos dados no repositório.

A seguir são apresentadas algumas ideias para compor a política de *upload* ou depósito, organizadas em seções.



#### 1. Dados de pesquisa

Sugere-se redigir um texto para informar o tipo de dado que pode ser armazenado no repositório, por exemplo, qualquer tipo de dados; ou especificamente, dados estatísticos, imagens, texto simples, texto estruturado, dados audiovisuais, dados bibliográficos, som, entre outros tipos.

Sugere-se que esta seção informe os formatos de arquivos aceitos, por exemplo, .csv, .txt, .sav; as licenças utilizadas pelo repositório, e o modelo de disponibilização dos dados, se total, imediata, embargada, entre outros modelos.

O repositório deve indicar necessidade de anonimização dos dados e controle / responsabilidade acerca de quaisquer informações sigilosas e sensíveis.

Recomenda-se a aplicação de licenças abertas, como as licenças *Creative Commons CC BY*, *Open Data Commons*, *AusGoal*, ou licenças específicas do próprio repositório. No entanto, a indicação do uso das *Creative Commons* é o mais comum e isso se deve ao fato de estas criarem instrumentos jurídicos flexíveis, de fácil compreensão de seu conteúdo, que se fundamentam nas prerrogativas que cada indivíduo tem como autor, de permitir o acesso e uso de suas obras (Lemos, 2005; Swan, 2016). A escolha da licença pode ficar a critério do autor dos dados, e também depende dos requisitos do financiador. Em alguns casos, a opção *CC0 Public Domain Dedication* (sem direitos reservados) pode ser mais apropriada.

Exemplos de redação:

O repositório permite o depósito de quaisquer tipos de dados de pesquisa, como por exemplo [citar exemplos, como dados estatísticos, entrevistas, dados observacionais] em formatos [informar se permite formatos proprietários ou somente formatos abertos]. O depósito de dados resulta em [sua publicação imediata, ou embargo permitido pelo tempo x, ou outra possibilidade]. Para submissão os dados devem ser [se for o caso, anonimizados ou submetidos a algum tipo de processo específico]. Os dados estarão disponíveis sob licença [informar a licença adotada pelo repositório, ou informar se a decisão acerca de licença fica a cargo do depositante].

## 2. Metadados

Nesta seção é preciso informar quem é responsável pela descrição dos metadados, se biblioteca, pesquisador, coordenador da comunidade, entre outras possibilidades, além dos metadados mínimos para descrição dos conjuntos de dados (conforme nível de especificidade de cada repositório). Também é importante informar a licença que se aplica aos metadados.

Os dados depositados devem ser descritos pelos metadados [autor, ano de criação, etc, informar todos os metadados obrigatórios] para que sejam

identificados e permitam reuso. O processo de descrição é de responsabilidade de [se biblioteca, pesquisador, coordenador da comunidade, entre outras possibilidades]. Os metadados dos itens depositados no Repositório [nome] são disponibilizados sob a licença [por exemplo, *Creative Commons Attribution CC BY (Attribution Only)*].

### 3. Submissão

Sugere-se descrever quem é responsável e tem permissão para realizar submissão, e uma breve descrição do processo. É importante mencionar a possibilidade de autodepósito. Por exemplo:

Os usuários autorizados a fazer *upload* de dados incluem [bibliotecários, pessoal, pesquisadores, professores, doutorandos, mestrandos, alunos de graduação, etc].

O depositante é responsável pela descrição de metadados [ou a biblioteca ou outro órgão o faz]. Metadados e itens a serem submetidos estão sujeitos a verificação pelos administradores do repositório, que examinam a elegibilidade de autores/depositantes, relevância para o escopo do Repositório, layout e formatos válidos.

Os depositantes devem possuir todos os direitos necessários para depositar os dados. Os depositantes de dados são responsáveis pelo cumprimento de quaisquer regulamentos legais nacionais ou internacionais e embargos de editores.

A validade e autenticidade do conteúdo das submissões e quaisquer violações de direitos autorais são de inteira responsabilidade do depositante.

Se o Repositório de Dados receber prova de violação de direitos autorais, o item relevante será removido o mais rápido possível.

Especificação de limite de tamanho dos arquivos, por exemplo: Arquivos com tamanho superior a [50] MB não podem ser carregados no repositório, exceto como permissão especial dos administradores do repositório.

Ao depositar os metadados e o conteúdo no Repositório de Dados, os depositantes aceitam os termos e condições definidos nesta Política do Repositório.

#### 4. Versionamentos

Informar se o Repositório opera com possibilidade de versionamento de arquivos, as chamadas versões efêmeras ou em andamento; ou aceita somente o depósito de arquivo final e consolidado, por exemplo:

O Repositório [nome] aceita depósito de [versão consolidada dos dados de pesquisa; ou, se necessário, uma versão atualizada poderá ser depositada em substituição à versão anterior, que será retirada ao público; ou, são permitidos quantos versionamentos o autor julgar necessário].

#### 5. Privacidade

Tendo em vista a LGPD, informar se o Repositório retém informações de acesso. Sugere-se que esta seção da política seja construída em conjunto com o setor de TI ou procuradoria da instituição. Algumas ideias para registrar nesta seção da política:

Sempre que um usuário visita o repositório, as seguintes informações são coletadas e armazenadas automaticamente [endereço de IP, tipo de navegador e sistema operacional, data e hora, páginas visitadas, localização, entre outros].

Essas informações são usadas internamente apenas para [solução de problemas técnicos, para melhorar a usabilidade do repositório e para registrar estatísticas agregadas]. O sistema coleta informações pessoais enviadas durante o processo de depósito. As informações enviadas serão usadas apenas para [o propósito para o qual você as enviou e não serão compartilhadas com terceiros ou usadas para fins comerciais].

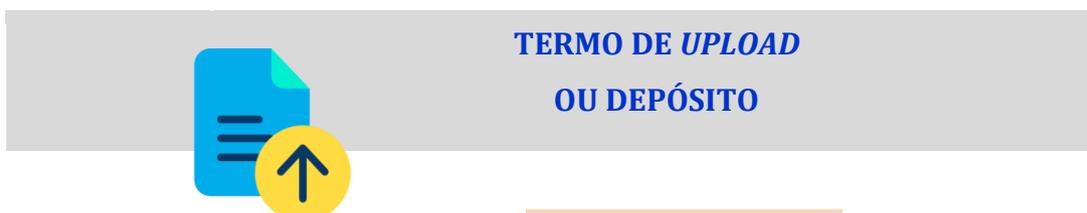
O Google Analytics é usado para analisar o uso deste repositório. A política de privacidade do Google está disponível em: <http://www.google.com/privacypolicy.html>.

[Se você usa aplicativos de terceiros, você deve listá-los na política de privacidade e fornecer links para suas políticas relacionadas a dados pessoais.]

Como visto, a política de *upload* ou depósito repete alguns trechos da política geral e inclui detalhes mais específicos, especialmente aqueles que podem ser alterados com maior frequência. Considerando que a política geral muitas vezes é aprovada pelo conselho deliberativo, reitoria ou direção das instituições, entende-se que é um documento mais perene. Há detalhes do processo de submissão que precisam ser alterados com maior frequência, como por exemplo, permissões, tamanho de arquivos, etc. Dessa forma, justifica-se a redação de uma política específica para *upload* ou depósito. Alguns repositórios optam por incluir este conteúdo em um documento único, denominado termo.

## 4.2 TERMO DE UPLOAD OU DEPÓSITO

Considerando que a política é um documento mais extenso e detalhado, muitos repositórios apresentam também o termo, que consiste em um documento mais sucinto. O aceite do termo pode ser solicitado a partir de um botão clicável. Há repositórios que entendem que usuário aceita os termos ao utilizar o serviço. Por exemplo:



Ao utilizar o serviço do repositório [nome do repositório], você declara que concorda e aceita os termos e condições aqui estabelecidos e as demais regras, políticas e procedimentos dos serviços, além das leis aplicáveis.

Um termo de *upload* ou depósito de dados de pesquisa é um documento que estabelece os direitos, responsabilidades e condições para o envio de conjuntos

de dados em um repositório de dados de pesquisa. Veja alguns exemplos e ideias para redação do termo:

Ao depositar no [nome do repositório] você reconhece que leu e concorda em sua totalidade com os Termos e condições de uso do repositório [nome] conforme declarado abaixo.

Você é o criador/autor ou recebeu permissão de um criador/autor para depositar qualquer conjunto de dados no [nome do repositório].

Para gerenciar e preservar adequadamente os conjuntos de dados para uso futuro, o [nome do repositório] requer permissões e garantias de você, o(s) criador(es), autor(es) e/ou o(s) proprietário(s) dos direitos autorais, incluindo a transferência de custódia de seus conjuntos de dados. Se a lei de direitos autorais for aplicável ao envio do conjunto de dados e você for o proprietário dos direitos autorais, ao aceitar este termo, você ainda mantém os direitos autorais de seu trabalho e não desiste do direito de enviar o conjunto de dados a editores ou outros repositórios. Se os direitos autorais forem aplicáveis e se você não for o proprietário dos direitos autorais, você declara que o proprietário dos direitos autorais lhe deu permissão para disponibilizar o conjunto de dados publicamente.

Ao depositar dados no [nome do repositório], você concede ao [nome do repositório ou instituição] o direito não exclusivo de reproduzir, traduzir (conforme descrito abaixo) e distribuir o conjunto de dados em qualquer formato ou meio em todo o mundo e livre de royalties, incluindo, mas não limitado, a publicação na Internet.

Você entende que os conjuntos de dados a serem publicados no [nome do repositório] devem estar de acordo com as instruções descritas para o depósito de dados de pesquisa. Você também entende que os dados depositados estão sujeitos à revisão e aprovação dos responsáveis pela curadoria antes de serem publicados.

Você concorda que [nome do repositório] pode converter os arquivos de dados e/ou arquivos de metadados depositados em qualquer meio ou formato e fazer

várias cópias do conjunto de dados depositado para fins de segurança, backup e preservação [se o repositório realizar essa prática]. Além disso, você entende e concorda que a equipe do [nome do repositório] pode, se necessário, fazer alterações nos metadados descritivos, incluindo aprimoramentos (como adição de palavras-chave) e/ou correções (como erros de digitação). Quaisquer alterações de metadados descritivos serão rastreadas por meio do controle de versão automatizado do [nome do repositório] [se for o caso]. Seu conjunto de dados não será alterado sem sua permissão.

Você entende que o nome do depositante é exibido com o conjunto de dados publicado, assim como os termos de uso selecionados pelo depositante.

Você entende que [nome do repositório] se compromete a preservar seu conjunto de dados publicado de acordo com a Política [de Preservação ou política geral do repositório]. O [nome do repositório] também se compromete a fornecer acesso público aos conjuntos de dados por meio da manutenção de páginas da web persistentes do [nome do repositório], registros de metadados descritivos e identificadores, e a fornecer métricas de acesso, que podem incluir visualizações de páginas, *downloads* e citações.

Se você não estiver de acordo com os itens acima e a seguir, você deverá entrar em contato com a equipe do [nome do repositório] através do [endereço de email] antes de depositar os dados no [nome do repositório].



## DECLARAÇÃO PARA DEPÓSITO DOS DADOS

Para depositar um conjunto de dados, você declara que:

- Você tem o direito de conceder os direitos contidos neste termo.
- Até onde você sabe, nada no conjunto de dados infringe os direitos autorais de alguém ou outros direitos de propriedade intelectual.
- Se o conjunto de dados contiver material do qual você não detém direitos autorais, você obteve a permissão irrestrita do proprietário dos direitos

autorais para conceder ao [nome do repositório] os direitos exigidos por este acordo e que esse material de propriedade de terceiros seja claramente identificado e reconhecido no conteúdo do conjunto de dados.

- Nada no conjunto de dados viola quaisquer termos (por exemplo, acordo de confidencialidade, acordo de transferência de material, termos de uso, etc.).
- Nada no conjunto de dados contém qualquer informação privada, informação confidencial, informação proprietária de terceiros, informação de exportação, controlada, dados protegidos ou informação que não deva ser compartilhada publicamente.
- Nada no conjunto de dados contém vírus de *software* ou códigos de computador, arquivos ou programas capazes de permitir acesso não autorizado ou interromper, danificar, limitar ou interferir no funcionamento adequado do [nome do repositório] ou *software*, hardware ou equipamento de telecomunicações de outros usuários.
- Você entende que, se os termos de direitos autorais ou propriedade do conjunto de dados forem alterados, é sua responsabilidade notificar o [nome do repositório] sobre essas alterações.
- Se o conjunto de dados for baseado em trabalho que foi patrocinado ou apoiado por uma agência ou organização, você deve cumprir todas as obrigações exigidas pelo contrato ou acordo com essa agência ou organização antes de depositar no [nome do repositório].

## MÓDULO V

### Objetivos

Discute as políticas relacionadas ao termo de *download*.

### Conceitos

Política de *Download*; Transferência de Dados; Termos de *Download*.

### 5.1 TERMO DE DOWNLOAD

Os repositórios de dados de pesquisa geralmente possuem termos de depósito e *download* que são estabelecidos para orientar os pesquisadores e usuários sobre como depositar e acessar os dados armazenados na plataforma. Esses termos têm como objetivo garantir a integridade dos dados, respeitar os direitos autorais e promover o uso ético e responsável das informações científicas.

O termo de *download* é baseado na política do repositório e deve conter informações sobre **licenças de uso** para definir as permissões e restrições associadas aos dados. Além de observar a **privacidade e confidencialidade** dos dados, incluindo diretrizes sobre o tratamento de dados sensíveis, como dados pessoais ou informações confidenciais. O objetivo do termo é a proteção de usuários e instituições do repositório, para garantir a segurança na gestão e no compartilhamento de informações científicas (LEITE, 2022, não paginado).

Apresentam-se a seguir algumas ideias para redação do termo de *download*:



## TERMO DE DOWNLOAD

Este documento estabelece os termos e condições para o *download* e uso de conjuntos de dados de pesquisa disponibilizados no [nome do repositório]. Ao fazer o *download* e utilizar esses dados, você concorda em cumprir os seguintes termos:

### 1. Regras de conduta

Como usuário, você pode fazer o *download* do conteúdo disponível publicamente no [nome do repositório].

Para registrar-se, você deve fornecer informações precisas, completas e atualizadas. O [nome do repositório] tem o direito de recusar o registro ou cancelar uma conta. O registro é feito através de [e-mail, identificador institucionais, ORCID].

Você é o único responsável pelas atividades que ocorrem em sua conta e será responsável por manter a confidencialidade da sua senha. Você nunca deve usar a conta de outro usuário sem permissão. Você notificará imediatamente o [nome do repositório] por escrito sobre qualquer uso não autorizado de sua conta, ou outra violação de segurança relacionada à conta.

Ao ter uma conta no [nome do repositório], você reconhece que:

1. os detalhes da sua conta (incluindo nome, sobrenome, endereço de e-mail e nome de usuário) são pesquisáveis por outros usuários;
2. os detalhes da sua conta podem ser gravados quando você faz o *download* de conjuntos de dados e essas informações podem ser visualizadas pelo autor;
3. seu primeiro nome, sobrenome e afiliação serão vinculados aos seus *uploads*.

## 2. Licenças de uso

O [nome do repositório] usa a licença [por exemplo, *Creative Commons CC BY*] como padrão. Essa opção permite que [outras pessoas redistribuam, remixem, adaptem e criem novos materiais a partir do seu trabalho, até mesmo comercialmente, desde que seja atribuído o devido crédito].

Os usuários do [nome do repositório] devem fomentar o bom uso dos dados relacionados a seres humanos. Os dados com essas características não devem ser utilizados para:

1. obter informações que permitam, direta ou indiretamente, identificar quaisquer sujeitos da pesquisa;
2. produzir e/ou publicar conexões entre conjuntos de dados que possam identificar indivíduos ou organizações;
3. obter informações (adicionais) sobre, ou ainda, meios de contato (adicionais) para sujeitos já identificados.

## 3. Citação e crédito à fonte dos dados

Ao utilizar os dados você deve fornecer a devida citação e atribuição aos pesquisadores responsáveis pela criação e disponibilização dos dados. As informações necessárias para a citação, como título do conjunto de dados, autores, identificador único do conjunto de dados e data de publicação, estão disponíveis na página de *download* dos dados.

## 4. Privacidade e Confidencialidade

Alguns conjuntos de dados podem conter informações sensíveis ou confidenciais. É sua responsabilidade garantir a privacidade e a proteção desses dados, cumprindo as regulamentações e leis aplicáveis [Lei x], bem como as diretrizes éticas de pesquisa [pode citar manual de ética institucional].

Você concorda em não divulgar ou identificar indivíduos ou instituições específicas sem a devida autorização, caso os dados contenham informações identificáveis.

## 5. Responsabilidades do usuário

Caso identifique erros, problemas ou inconsistências nos dados, você é encorajado a entrar em contato com os responsáveis pelo repositório para fornecer *feedback* e contribuir para a melhoria contínua da qualidade dos dados.

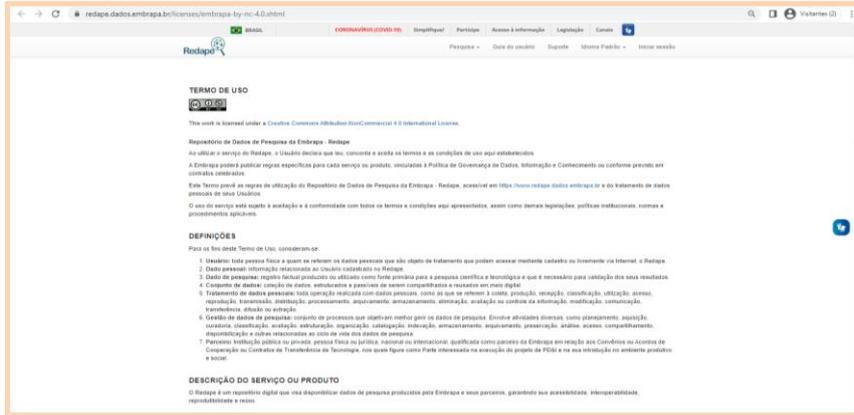
## 6. Restrições de uso

Alguns conjuntos de dados podem estar sujeitos a restrições adicionais de uso, como limitações de acesso por período determinado ou restrições geográficas. É sua responsabilidade cumprir essas restrições ao utilizar os dados.

Ao fazer o *download* dos dados, você confirma que leu, entendeu e concorda com os termos deste documento. Este Termo de *Download* de Dados de Pesquisa é válido a partir da data de *download* e permanece em vigor até que você exclua ou deixe de utilizar os dados baixados.

EXEMPLOS DE TERMOS:

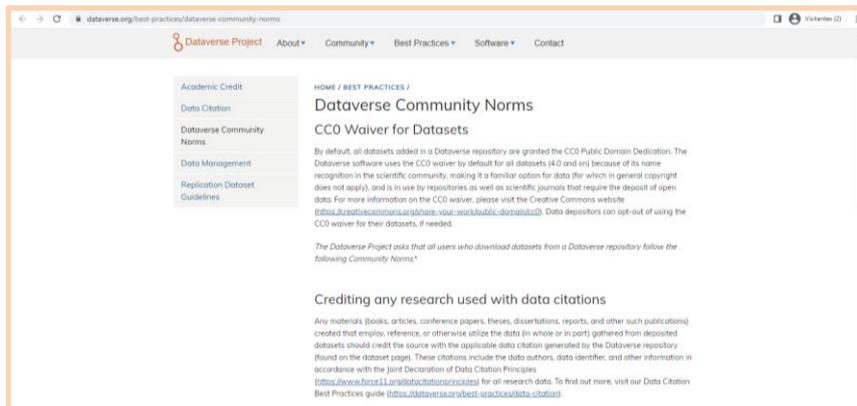
**Repositório de Dados de Pesquisa da Embrapa (Redape)**



**Repositório de Dados de Pesquisa da Unicamp (REDU)**



**Dataverse Community Norms**



## MÓDULO VI

### Objetivos

Apresenta as possibilidades que os fluxogramas oferecem para mapear os passos necessários para *upload* e *download* de dados de pesquisa em repositórios destinados a este fim.

### Conceitos

Fluxogramas; *Upload* de Dados; *Download* de Dados.

### 6.1 FLUXOGRAMAS

Um fluxograma é uma representação gráfica que ilustra a sequência de etapas, processos ou atividades de um sistema, processo ou projeto. Ele usa símbolos padronizados e setas para mostrar a direção do fluxo e a conexão entre as diferentes etapas. No contexto dos repositórios de dados abertos, os fluxogramas são fundamentais para o adequado funcionamento do repositório, considerando-se a quantidade de usuários envolvidos. A possibilidade de auto depósito também demanda a publicização dos fluxogramas, para evitar que o usuário pule etapas dos processos de depósito ou *download* de arquivos.

## 6.2 EXEMPLO DE FLUXOGRAMA

**Veja um exemplo de fluxograma para *upload* de dados de pesquisa em um repositório (SciELO Data):**

Este fluxograma descreve o passo-a-passo para o *upload*/depósito no

<https://data.scielo.org/>.

Para realizar o *upload* do seu conjunto de dados, siga as seguintes etapas:



01 Acesse a página do SciELO

Data: <https://data.scielo.org/> e



02 Faça login na sua **conta**

**CONTA:** um meio de acesso controlado pelo usuário ao Repositório que é protegida por um nome de usuário (login) e senha, criados pelo usuário, com o qual os usuários podem obter acesso às funcionalidades de *upload* e *download*. (PAVÃO, 2019).



03 Se você ainda não possui uma

conta, clique em **[registrar ou criar conta]**



04 Após fazer *login*, clique em adicionar dados



05 Preencha os metadados de citação: será exibida uma página onde você deverá fornecer as informações relevantes sobre o depositante.



**DEPOSITANTE:** Quem faz *upload* de uma coleção de arquivos de dados como um conjunto de dados para um *dataverse* no aplicativo (PAVÃO, 2019).

**INFORMAÇÕES DE REGISTRO:** informações necessárias para registrar uma conta, incluindo, entre outras, nome, e-mail, instituição, cargo e número de telefone (PAVÃO, 2019).

**METADADOS:** informações sobre um item no repositório, incluindo, mas não limitado, ao nome do autor, data de publicação, título do conteúdo de dados, descrição do conteúdo e outras informações relacionadas. Deve estar em conformidade com um padrão para permitir a interoperabilidade entre computadores (PAVÃO, 2019).

06 Faça o *upload* do arquivo com o conjunto de dados



**CONJUNTO DE DADOS:** um conjunto de arquivos contendo dados de pesquisa e documentação necessária para tornar os dados reutilizáveis (PAVÃO, 2019).

07 Informe o tipo de licença



08 Se o conjunto de dados possuir restrição de acesso: escolha o tipo de restrição



09 Se o conjunto de dados não possuir restrição de acesso: prossegue



10 Salve as alterações



11 Submeta para revisão



**ENVIOS DO USUÁRIO:** *Dataverses* e conjuntos de dados carregados (*uploaded*) (PAVÃO, 2019).

**DOCUMENTAÇÃO:** quaisquer arquivos digitais, como um *codebook*, relatório técnico ou metodológico ou guia do usuário, que expliquem a produção, proveniência, processamento ou interpretação dos dados da pesquisa (PAVÃO, 2019).

12



## Veja um exemplo de fluxograma para *download* de dados de pesquisa em um repositório (*SciELO Data*):

Este fluxograma descreve o passo-a-passo para o *download* de dados no <https://data.scielo.org/>.

Para realizar o *download* de um conjunto de dados, siga as seguintes etapas:



01 Acesse a página do *SciELO Data*: <https://data.scielo.org/> e



02 Na caixa de busca encontre o conjunto de dados de seu interesse



03 Clique no título



04 Na página do conjunto de dados, leia o Termo de Uso



- 05 Clique em *download* para baixar o documento *Word* com informações sobre os dados e o conjunto de dados



- 06 Não esqueça de citar corretamente o conjunto de dados

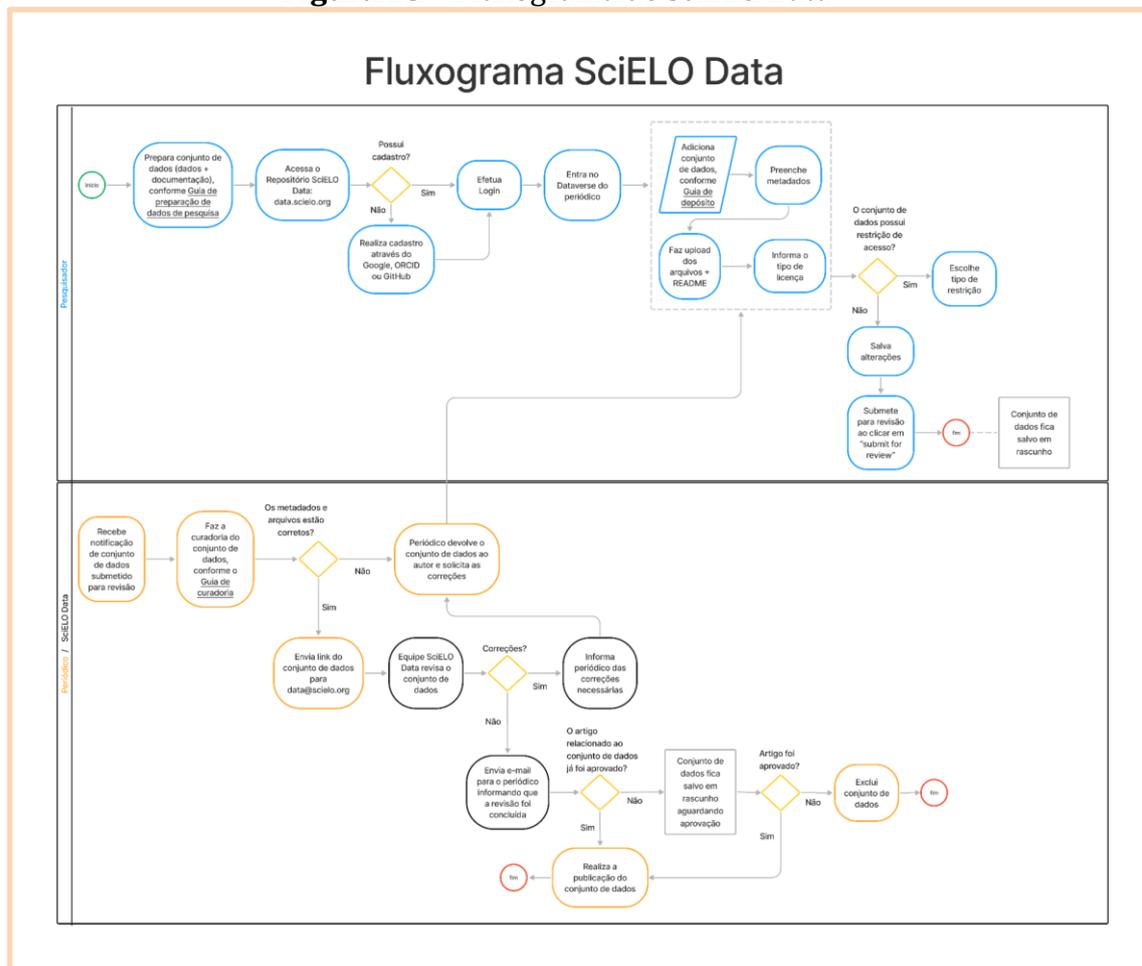


- 07



## Fluxograma do repositório de dados *SciELO Data*:

Figura 13 – Fluxograma do *SciELO Data*



Fonte: *SciELO Data* (2023).

## REFERÊNCIAS

BATISTA, M. C. P. Organização, sistemas e métodos em bibliotecas do Senac no Amazonas. Senac. **DOC**: revista de informação e conhecimento, v. 3, n. 1, p. 18-31, 2016. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/163751>. Acesso em: 14 jun. 2023.

BORGMAN, C. L. Research Data: who will share what, with whom, when an why. (RatSWDWorking Paper n. 161. Oct. 2010).

CARVALHO, C. P.; CARVALHO, R. A. Construção de políticas para repositórios institucionais: análise da ferramenta do OpenDOAR. **Biblos**: Revista do Instituto de Ciências Humanas e da Informação, v. 26, n. 2, p. 105-138, 2012. Recuperado de <https://periodicos.furg.br/biblos/article/download/3450/2211/10380>.

PAVÃO, C. G., et al. **Acesso aberto a dados de pesquisa no Brasil**: políticas para repositórios de dados de pesquisa. 2019. Disponível em: <http://hdl.handle.net/20.500.11959/1263>. Acesso em: 26 set. 2023.

SAYÃO, L. F. O papel dos repositórios digitais na curadoria de dados de pesquisa. In: VECHIATO, Fernando *et al.* (org.). **Repositórios digitais**: teoria e prática. Curitiba: EDUTFPR, 2017. Disponível em: <http://repositorio.utfpr.edu.br/jspui/handle/1/2495>. Acesso em: 17 maio 2023.

SILVEIRA, J. I. **Acesso aberto a dados de pesquisa em repositórios universitários internacionais**: um estudo sobre políticas de depósito, acesso e uso. 2021. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação da Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2021. Disponível em: <https://lume.ufrgs.br/handle/10183/222095>. Acesso em: 27 set. 2023.

SILVEIRA, J. I.; VANZ, S. A. S. Diretrizes para políticas de depósito, acesso e uso de dados de pesquisa: proposta a partir da análise de repositórios de dados universitários internacionais. **AtoZ**, 2023, no prelo.

## FONTES CONSULTADAS

DIGITAL CURATION CENTRE. **Digital Curation**. Edinburgh: DCC, 2013. Disponível em: <https://www.dcc.ac.uk>. Acesso em: 20 jun. 2023.

GABRIEL JUNIOR, R. F., *et al.* **Acesso aberto a dados de pesquisa no Brasil**: dataverse: documentação de instalação e configuração – Ubuntu, 2021. Disponível em: <https://cedap.ufrgs.br/jspui/handle/20.500.11959/1202>. Acesso em: 26 set. 2023.

GABRIEL JUNIOR, Rene Faustino *et al.* Acesso aberto a dados de pesquisa no Brasil: mapeamento de repositórios, práticas e percepções dos pesquisadores e tecnologias. **Ciência da Informação**, v. 48, p. 87-101, 2019. Disponível em: <http://hdl.handle.net/10183/212266>. Acesso em: 26 set. 2023.

MELIS, M. F. M. S. **Acesso aberto aos dados de pesquisa nas universidades brasileiras e os indicadores de CT&I**. 2018. Dissertação (Mestrado

Profissional em Propriedade Intelectual e Transferência de Tecnologia para a Inovação) – Programa de Pós-Graduação em Propriedade Intelectual  
Transparência de Tecnologia para a Inovação, Universidade de Brasília, Brasília, 2018. Disponível em: <http://repositorio.unb.br/handle/10482/34537>. Acesso em: 26 set. 2023.

MONTEIRO, E. C. S. A. **Direitos autorais nos repositórios de dados científicos**: análise sobre os Planos de Gerenciamento dos Dados. 2017. Dissertação (Mestrado em Informação e Tecnologia) – Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2017. Disponível em: <https://repositorio.unesp.br/handle/11449/149748>. Acesso em: 26 set. 2023.

PAVÃO, C. M. G. *et al.* Proposta de criação de uma rede de dados abertos da pesquisa brasileira. Revista Digital de Biblioteconomia e Ciência da Informação, v. 16, p. 329-343, 2018. Disponível em: <http://hdl.handle.net/10183/175284>.

SANCHEZ, F. A.; VIDOTTI, S. A. B. G.; VECHIATO, F. L. A contribuição da curadoria digital em repositórios digitais. **Revista Informação na Sociedade Contemporânea**, Natal, número especial, p. 11-17, 2017. Disponível em: <https://periodicos.ufrn.br/informacao/article/download/12280/8508>. Acesso em: 26 set. 2023.

SAYÃO, L. F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, Londrina, v. 1, n. 21, p. 90-115, maio/ago., 2016a. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/download/27939/20122>. Acesso em: 16 maio 2023.

SAYÃO, L. F.; SALES, L. F. Curadoria digital de dados de pesquisa: desafios informacionais, gerenciais, tecnológicos e políticos. In: SEMINÁRIO EM CIÊNCIA DA INFORMAÇÃO, 6. Londrina, 2016b. Slides...Disponível em: <https://brapci.inf.br/index.php/res/download/156514>. Acesso em: 26 set. 2023.

UNIVERSIDADE FEDERAL DO CEARÁ. **Manual Técnico de Elaboração de Manuais Administrativos**. Fortaleza: UFC, 1999.

UNIVERSIDADE FEDERAL DE MARINGÁ. Pró-Reitoria de Planejamento e Desenvolvimento Institucional. **Regulamento**. Maringá, 2023.

VEIGA, V.; ALVES, A. S. **Repositórios**: conceito, tecnologia e aplicação. Rio de Janeiro: FIOCRUZ/ICICT, 2016. 76 p. Disponível em: <https://www.arca.fiocruz.br/handle/icict/16385>. Acesso em: 17 maio. 2023.